

TABLE S4: Comparison of LD estimates by the proposed ML estimator to those by an imputation-based method when major allele frequencies at two sites of interest, p and q , are high.

D	μ	(p, q)	Φ	Method	\hat{D} (mean \pm SD)	\hat{D}^2 (mean \pm SD)	Asymptotic theoretical SD(\hat{D})	Theoretical $E[\hat{D}^2]$	RMSD(\hat{D})	RMSD(\hat{D}^2)
0.01	2	(0.9,0.9)	0	ML	0.012 \pm 0.014	0.00034 \pm 0.00049	0.011	0.00021	0.014	0.00054
0.01	2	(0.9,0.9)	0	Imputation	0.003 \pm 0.005	0.00003 \pm 0.00005	0.011	0.00021	0.009	0.00008
0.01	2	(0.9,0.9)	1/3	ML	0.011 \pm 0.014	0.00030 \pm 0.00043	0.008	0.00017	0.014	0.00047
0.01	2	(0.9,0.9)	1/3	Imputation	0.003 \pm 0.006	0.00005 \pm 0.00015	0.008	0.00017	0.009	0.00016
0.01	2	(0.9,0.9)	1	ML	0.011 \pm 0.011	0.00022 \pm 0.00031	0.008	0.00017	0.011	0.00033
0.01	2	(0.9,0.9)	1	Imputation	0.004 \pm 0.006	0.00005 \pm 0.00009	0.008	0.00017	0.009	0.00010
0.01	10	(0.9,0.9)	0	ML	0.009 \pm 0.010	0.00017 \pm 0.00024	0.011	0.00021	0.010	0.00025
0.01	10	(0.9,0.9)	0	Imputation	0.005 \pm 0.008	0.00008 \pm 0.00013	0.011	0.00021	0.010	0.00013
0.01	10	(0.9,0.9)	1/3	ML	0.010 \pm 0.009	0.00018 \pm 0.00023	0.008	0.00017	0.009	0.00024
0.01	10	(0.9,0.9)	1/3	Imputation	0.006 \pm 0.008	0.00009 \pm 0.00013	0.008	0.00017	0.009	0.00013
0.01	10	(0.9,0.9)	1	ML	0.010 \pm 0.008	0.00016 \pm 0.00018	0.008	0.00017	0.008	0.00019
0.01	10	(0.9,0.9)	1	Imputation	0.006 \pm 0.008	0.00011 \pm 0.00025	0.008	0.00017	0.009	0.00025
0.05	2	(0.9,0.9)	0	ML	0.048 \pm 0.021	0.00280 \pm 0.00245	0.016	0.00271	0.021	0.00245
0.05	2	(0.9,0.9)	0	Imputation	0.011 \pm 0.010	0.00022 \pm 0.00044	0.016	0.00271	0.041	0.00232
0.05	2	(0.9,0.9)	1/3	ML	0.048 \pm 0.020	0.00270 \pm 0.00204	0.014	0.00266	0.020	0.00204
0.05	2	(0.9,0.9)	1/3	Imputation	0.015 \pm 0.013	0.00039 \pm 0.00054	0.014	0.00266	0.037	0.00218
0.05	2	(0.9,0.9)	1	ML	0.049 \pm 0.018	0.00274 \pm 0.00204	0.014	0.00266	0.018	0.00204
0.05	2	(0.9,0.9)	1	Imputation	0.028 \pm 0.017	0.00106 \pm 0.00104	0.014	0.00266	0.027	0.00177
0.05	10	(0.9,0.9)	0	ML	0.051 \pm 0.015	0.00284 \pm 0.00153	0.016	0.00271	0.015	0.00156
0.05	10	(0.9,0.9)	0	Imputation	0.046 \pm 0.022	0.00257 \pm 0.00189	0.016	0.00271	0.023	0.00188
0.05	10	(0.9,0.9)	1/3	ML	0.051 \pm 0.015	0.00282 \pm 0.00158	0.014	0.00266	0.015	0.00161
0.05	10	(0.9,0.9)	1/3	Imputation	0.044 \pm 0.022	0.00242 \pm 0.00189	0.014	0.00266	0.023	0.00188

0.05	10	(0.9,0.9)	1	ML	0.052±0.014	0.00284±0.00152	0.014	0.00266	0.014	0.00156
0.05	10	(0.9,0.9)	1	Imputation	0.045±0.021	0.00248±0.00186	0.014	0.00266	0.021	0.00185

Sample means and standard deviations of the LD coefficient \hat{D} and its square \hat{D}^2 estimated from simulated data by the ML estimator and the imputation-based method are compared with different values of the parameters. Root mean square deviations (RMSD) of the LD estimates are also compared. The comparisons are made when the mean depth of coverage, μ is low (2) or moderately high (10). Furthermore, comparisons with different values of the probability that an informative read covers both polymorphic sites, ϕ , are made. The theoretical prediction of the asymptotic standard deviation, which is calculated as a square root of the sampling variance given by Equation 9 (when $\phi > 0$) or 10 (when $\phi = 0$), is the expected level of achievement when individual genotypes are known without errors. Sample size $N = 100$, error rate $\varepsilon = 0.01$. A total of 100 simulation replicates were run for each set of parameter values.